Descriptors for Amino Acids Using MolSurf Parametrization

ULF NORINDER, PETER SVENSSON
Astra Pain Control AB, S-151 85 Södertälje, Sweden

Received 17 March 1997; accepted 12 August 1997

ABSTRACT: This work describes a new set of amino acid descriptors based on *ab initio* quantum mechanical calculations and MolSurf technology. These descriptors have been applied to two dipeptide data sets using partial least squares as the statistical engine. Statistically significant models for both data sets have been developed. The results from the derived peptide QSAR models are easy to interpret in terms of the theoretically computed MolSurf parameters of physicochemical nature. © 1998 John Wiley & Sons, Inc. *J Comput Chem* **19**: 51–59, 1998

Keywords: MolSurf; QSAR; amino acid descriptors; quantum chemistry; PLS

Introduction

In quantitative structure–activity relationship (QSAR) studies it is important to be able to describe the compounds under investigation not only in a statistically correct manner but also, equally importantly, in a way that allows a straightforward interpretation in terms of physicochemical properties that are important for biological activity.

Traditionally, there are several ways to describe structures using tabulated parameters[1, 2] (e.g., σ_m and σ_p, Hansch's π, Swain–Lupton's F and R, etc.) as well as calculated properties at some appropriate quantum mechanical (QM) level of approximation, (e.g., HOMO, LUMO, atomic charges, and delocalization).

Because the computed descriptors (see Method of Calculation section and Table I) are relatively easy for medicinal chemists to interpret in terms of physicochemical properties, we wanted in this work to explore the conformational stability and emphasize the interpretability of some new QM derived descriptors using MolSurf technology.[3] We also wanted to demonstrate how these computed descriptors from theoretically calculated surface properties can be used to construct QSARs with good statistical significance.

We used the 20 natural amino acids as our data set for several reasons.

1. These amino acids contain a variety of functional groups, also found in many pharma-

Correspondence to: Dr. Ulf Norinder; e-mail: Ulf.Norinder@pain.se.astra.com

CCC 0192-8651 / 98 / 010051-09

TABLE I.
Computed MolSurf Descriptors.

Descriptor	Designation
Charge transfer for carbons	CT
Lewis base	LB
Lewis acid	LA
Maximum electrostatic potential	V_{max}
Minimum electrostatic potential	V_{min}
Maximum local ionization energy	I_{max}
Minimum local ionization energy	I_{min}
Octanol / water partition coefficient	log P
Polarity	
pK_a for nitrogen bases	pK_a
Polarizability	
Surface area	

ceutical drugs, that are important for biological activity.

2. The amino acids are rather flexible, which coincides with our intention to investigate the conformational stability of the computed descriptors.
3. There are several peptide data sets available that have previously been studied using other parametrizations that allow for comparisons with our results.

Method of Calculation

A summary of the computational protocol described below is depicted in Figure 1.

CONFORMATIONAL ANALYSIS

The input structures of the 20 natural (coded) amino acids for the conformational analysis were obtained from the library of amino acids available in the Spartan program package.[4] The amino acids were modeled in their neutral forms. Conformational analyses were performed with the Monte Carlo procedure in MacroModel[5] using the MM2 force field. One hundred starting conformations were generated for each amino acid. Unique mini-

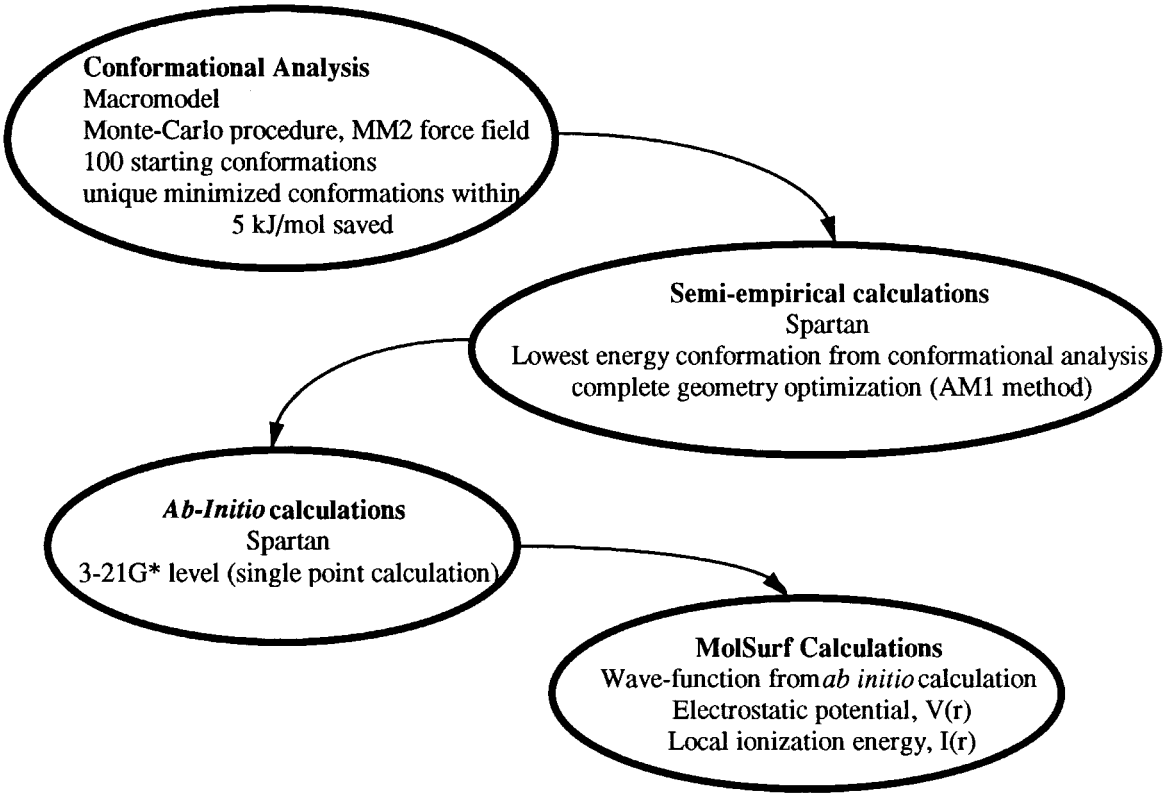

FIGURE 1. A summary of the computational protocol described in the Method of Calculation section.

mized conformations within 5 kJ/mol of the conformation with lowest energy found were saved for further studies.

SEMIEMPIRICAL CALCULATIONS

Each of the saved conformations from the previous conformational analyses was subjected to a complete geometry optimization (energy minimization) using the AM1 method (Hamiltonian) in Spartan.[4]

AB-INITIO CALCULATION

An *ab initio* calculation at the 3−21G* level without further geometry optimization (single point calculation) was performed on all AM1 optimized conformations using Spartan.[4]

MolSurf CALCULATIONS

The wave-function from each *ab initio* calculation was used by MolSurf[3] to compute various properties related to the molecular valence region that is represented by a surface of constant electron density (0.001 electrons/bohr^3) encompassing the molecule. The electrostatic potential, $V(r)$, and the local ionization energy, $I(r)$, are calculated at points evenly distributed (0.28 bohr apart) on this surface. Each point on the surface is then assigned to the closest atom of the investigated molecule. Thus, a number of points and, hence, a number of electrostatic potentials as well as local ionization energies are in this way associated with each atom and used to characterize its properties. Politzer, Sjöberg, Brinck, Haberlein, and coworkers demonstrated that these calculated surface properties can be used to construct statistical functions with good correlations to a wide spectrum of physicochemical properties.[6−12] Hydrophobicity[6−8] as well as pK_a,[9] hydrogen bonding,[9,10] and polarizability[7,11,12] are examples of such correlations used by MolSurf. For a recent review on MolSurf methodology, see ref. 13.

In this work MolSurf parameters were calculated for the entire amino acid as well as for the side-chain and individual atoms (i.e., the nitrogen atom of the amino group and the α-carbon atom).

PRINCIPAL COMPONENT ANALYSIS (PCA)

A PCA[14−16] was performed on the MolSurf descriptor matrix of the 20 coded amino acids. The data matrix was mean centered and autoscaled prior to the statistical operations.

PARTIAL LEAST SQUARES (PLS) ANALYSIS

The PLS method[17] was used to study two dipeptide data sets. The first data set, which consisted of 58 dipeptides related to ACE inhibition, were part of the development of the antihypertensive drug captopril.[18] The second data set was a compilation of 48 bitter tasting dipeptides.[18] Each of the two amino acid residue positions in the dipeptides was characterized by the linear and squared terms of the 22 calculated MolSurf parameters (see Tables II and III for a list of parameters and values, respectively). Thus, each dipeptide was described with 88 MolSurf variables. The data matrix for each data set was mean centered and autoscaled prior to the statistical analysis. Cross-validation was performed using a leave one out (LOO) methodology.[19]

TABLE II.
Correlation Coefficients between MolSurf Descriptors for Lowest Energy Conformer and Boltzmann Distributed Average of Retained Low Energy Conformations.

Property[a]	Substructure[b]	Corr. Coeff.
Surface	Side chain	0.9999
Polarizability	Side chain	0.9999
Surface	Molecule	0.9998
I_{min}	Atom N (amino group)	0.9998
Polarizability	Molecule	0.9997
I_{max}	Atom N (amino group)	0.9994
I_{max}	Atom C (α-carbon)	0.9994
log P	Molecule	0.9991
log P	Side-chain	0.9989
I_{min}	Atom C (α-carbon)	0.9989
Polarity	Molecule	0.9968
Polarity	Side chain	0.9964
V_{min}	Atom N (amino group)	0.9959
LA	Molecule	0.9948
LB	Molecule	0.9888
pK_a	Atom N (amino group)	0.9860
LA	Side chain	0.9857
V_{max}	Atom N (amino group)	0.9789
LB	Side chain	0.9614
V_{max}	Atom C (α-carbon)	0.9488
V_{min}	Atom C (α-carbon)	0.9485
CT	Atom C (α-carbon)	0.9038

[a] For a description of the MolSurf properties, see Table I.
[b] Molecule, the entire amino acid in its neutral form; amino group, the amino group attached to the α-carbon.

TABLE III.
Computed MolSurf Descriptors for 20 Coded Amino Acids.

	1 V_{min} N	2 V_{max} N	3 I_{min} N	4 I_{max} N	5 pK_a N	6 V_{min} αC	7 V_{max} αC	8 I_{min} αC	9 I_{max} αC	10 CT αC
Ala	−47.427	25.895	−0.430	−0.708	8.095	−30.767	28.045	−0.561	−0.743	−2.644
Arg	−44.933	25.115	−0.436	−0.714	7.147	−27.758	30.297	−0.569	−0.713	−1.969
Asn	−38.725	28.492	−0.450	−0.720	5.196	−16.326	32.051	−0.575	−0.738	−4.028
Asp	−41.817	20.034	−0.437	−0.715	7.008	−28.807	24.856	−0.566	−0.744	−3.322
Cys	−40.684	28.519	−0.445	−0.701	5.876	−24.929	36.075	−0.459	−0.731	−2.837
Gln	−43.968	18.338	−0.439	−0.716	6.673	−29.471	32.462	−0.546	−0.737	−2.934
Glu	−38.677	17.585	−0.441	−0.715	6.389	−20.348	22.057	−0.580	−0.748	−3.779
Gly	−45.272	26.781	−0.438	−0.726	6.789	−24.925	33.620	−0.561	−0.759	−3.606
His	−47.134	21.156	−0.432	−0.686	7.695	−29.818	21.510	−0.504	−0.723	−1.662
Ile	−46.949	24.836	−0.435	−0.688	7.421	−18.100	25.003	−0.554	−0.726	−1.881
Leu	−46.288	26.079	−0.435	−0.700	7.406	−19.630	29.344	−0.548	−0.736	−3.649
Lys	−42.745	20.469	−0.442	−0.713	6.415	−25.582	27.290	−0.570	−0.750	−3.412
Met	−38.920	23.648	−0.448	−0.719	5.583	−30.625	30.620	−0.526	−0.759	−3.694
Phe	−43.888	20.816	−0.435	−0.692	7.291	−29.901	20.033	−0.492	−0.729	−1.563
Pro	−44.254	11.816	−0.431	−0.709	7.811	−32.026	12.404	−0.564	−0.777	−2.734
Ser	−43.559	25.417	−0.440	−0.724	6.616	−28.065	26.081	−0.552	−0.740	−3.029
Thr	−43.967	22.944	−0.440	−0.722	6.602	−25.206	24.962	−0.558	−0.734	−2.734
Trp	−45.131	20.079	−0.430	−0.687	7.877	−29.553	18.090	−0.492	−0.725	−1.292
Tyr	−43.217	23.203	−0.436	−0.693	7.207	−24.400	19.868	−0.490	−0.729	−1.668
Val	−43.874	21.386	−0.442	−0.712	6.340	−26.173	32.336	−0.553	−0.718	−0.815

	11 Surface m	12 log P m	13 Polarizability m	14 Polarity m	15 LB m	16 LA m	17 Surface s	18 log P s	19 Polarizability s	20 Polarity s	21 LB s	22 LA s
Ala	124.669	1.103	0.952	0.559	0.725	1.261	37.988	−0.093	0.168	0.227	0.507	0.311
Arg	214.078	3.309	2.189	0.584	1.160	1.576	135.583	1.333	1.135	0.552	1.160	0.687
Asn	156.011	1.456	1.305	0.644	1.052	1.759	72.969	−0.753	0.432	0.665	1.052	0.595
Asp	152.513	1.850	1.242	0.620	0.580	1.532	68.631	−0.220	0.372	0.589	0.453	1.511
Cys	148.596	1.893	1.281	0.534	0.534	1.553	65.063	0.477	0.407	0.343	0.394	0.540
Gln	178.958	2.122	1.623	0.647	1.015	1.501	95.356	−0.344	0.641	0.630	1.015	0.960
Glu	170.073	2.371	1.484	0.605	0.647	1.425	87.730	0.384	0.551	0.557	0.559	1.299
Gly	105.640	0.592	0.725	0.670	0.734	1.414	7.206	−0.616	0.014	0.163	0.321	0.112
His	183.436	2.397	1.732	0.603	0.867	1.473	100.603	0.249	0.722	0.566	0.867	1.473
Ile	177.699	2.608	1.672	0.415	0.711	1.311	98.084	1.529	0.718	0.189	0.651	0.309
Leu	182.107	2.676	1.735	0.383	0.691	1.381	102.422	1.678	0.768	0.153	0.687	0.284
Lys	199.806	2.980	1.991	0.476	0.886	1.377	116.134	1.092	0.916	0.316	0.886	0.316
Met	191.341	3.043	1.891	0.481	0.702	1.442	107.039	1.581	0.840	0.314	0.466	0.414
Phe	204.284	3.428	2.064	0.480	0.736	1.298	121.451	2.021	0.976	0.327	0.473	0.199
Pro	153.003	1.788	1.305	0.457	0.736	1.426	77.446	1.005	0.481	0.149	0.405	0.171
Ser	131.385	1.323	1.003	0.601	0.766	1.677	44.705	−0.733	0.204	0.555	0.766	0.360
Thr	149.855	1.772	1.246	0.532	0.739	1.641	67.162	−0.043	0.389	0.405	0.739	0.366
Trp	236.955	4.108	2.587	0.536	0.785	1.344	153.493	2.379	1.379	0.454	0.571	1.344
Tyr	212.749	3.397	2.159	0.557	0.730	1.632	128.937	1.547	1.038	0.472	0.516	1.632
Val	160.979	1.997	1.429	0.437	0.713	1.400	79.965	1.117	0.521	0.165	0.494	0.216

See Table II for an explanation of the MolSurf variables: m, the entire molecule (amino acid); s, amino acid side chain; αC, α-carbon, N, nitrogen atom of the amino group attached to the α-carbon atom.

A PLS analysis was also performed to investigate the relations between the z scales of Hellberg et al.[20] and the 11 MolSurf parameters reported in this work.

Results and Discussion

MolSurf DESCRIPTORS

The Boltzmann distributed average was computed for each of the derived MolSurf parameters for every amino acid. In general, the computed MolSurf descriptors are rather insensitive to low energy conformational variability of the amino acid. The correlations of the MolSurf parameters between the Boltzmann weighted average values and the corresponding values for the lowest energy conformation are quite satisfactory and have squared correlation coefficients of between 0.904 and 0.999 (see Table II for details of each parameter). However, for some of the side-chain properties of amino acids having small side chains, as well as some electronic properties related to the computed maximum and minimum electrostatic potentials for the nitrogen atom of the amino group attached to the α-carbon atom and the α-carbon atom itself, somewhat larger variations in relation to the Boltzmann distributed average was observed (see Fig. 2). The former observation relates to values such as log P (Ala, Thr, and Gln), and Lewis base (Gly, Ala, and Leu), which all have rather small numerical values for these parameters. (The parameters are available from the authors (U.N.) on request.) However, in these cases the absolute difference between the Boltzmann distributed average and the corresponding value for the lowest energy conformation is rather small compared to the range of values for the property in question (log P and Lewis base, respectively) exhibited by the 20 natural amino acids. Somewhat more care has to be taken for the latter observation when using the values related to properties of

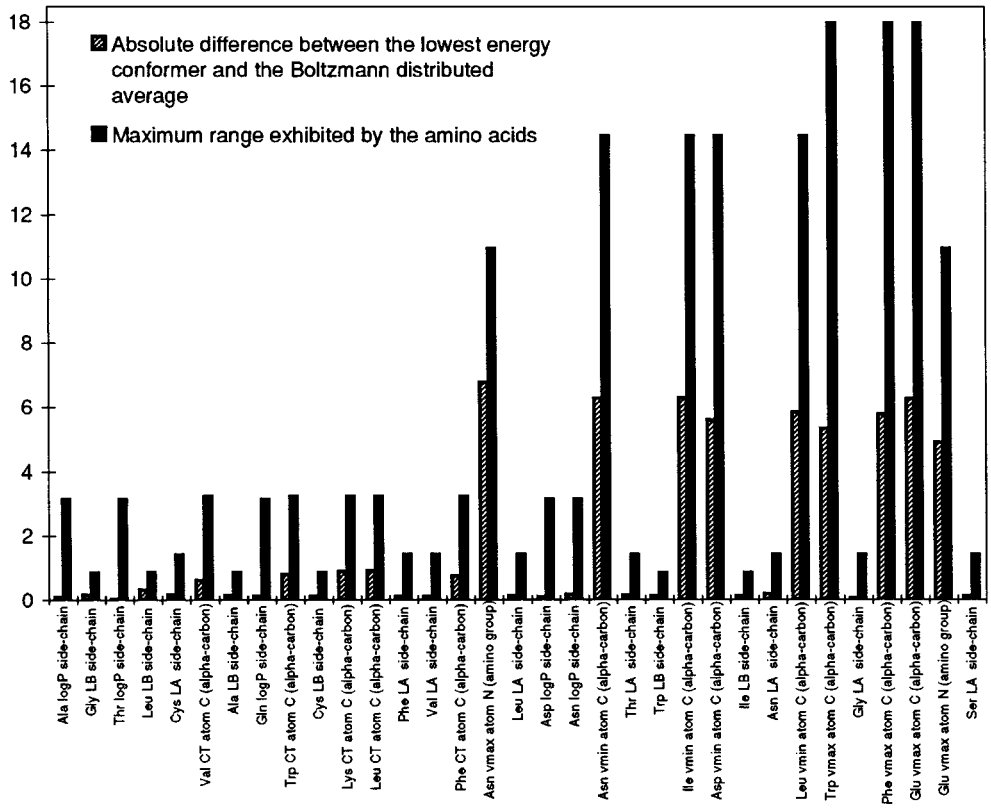

FIGURE 2. Bar chart of the absolute difference between the lowest energy conformer and the Boltzmann distributed average for the computed MolSurf descriptors. The largest absolute differences in relation to the average values (> 0.20) are depicted in descending order.

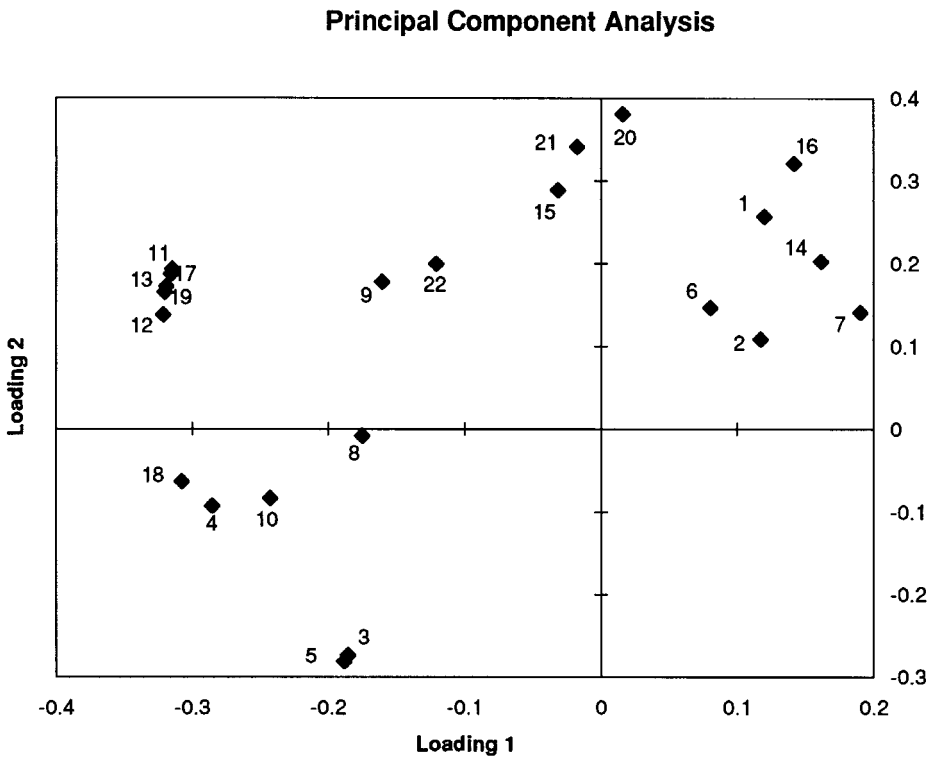

FIGURE 3. The first two loadings from the principal component analysis of the MolSurf descriptors for the 20 coded amino acids.

electrostatic nature (minimum and maximum potentials, respectively) of individual atoms. Here the variations in relation to the range of values found in the 20 natural amino acids are somewhat larger. The relative insensitive behavior to conformational variability for the calculated MolSurf descriptors can probably be associated with two reasons. The first is that we were interested in investigating the variability of low energy (≤ 5 kJ/mol of the lowest found conformation) conformers. Thus, both of the computed properties (electrostatic potential and local ionization energy), which are calculated from the resulting wave functions of low energy conformers, show relatively little variance across these sampled conformations. By using high energy conformations (which would not contribute to the Boltzmann average), it is possible to obtain quite different values for the MolSurf descriptors.[21] However, this was not the objective of this investigation (see above).

The second reason is that the surface used to calculate the MolSurf descriptors is larger and smoother than the corresponding van der Waals surface around the molecule, and a large number of surface points are associated with each atom of the molecule. Both of the computed properties (electrostatic potential and local ionization energy) form somewhat of a continuum when moving from one point on the surface to the next. This situation is rather similar to the one occurring in grid and surface based 3D-QSAR applications[22] where moving from one grid or surface point to another does not dramatically change the electrostatic interactions (potential) exhibited by the probe. Yet, useful models that explain pK_a values[23] and lipophilicity[24] have been established. Because the calculated MolSurf descriptors (see Table I) are derived from the electrostatic potential and the local ionization energy and extreme points (i.e., maximum and minimum), thereof the variation from one low energy conformer to another is not that significant. Thus, the derived MolSurf parameters are not sensitive to low energy conformational behavior.

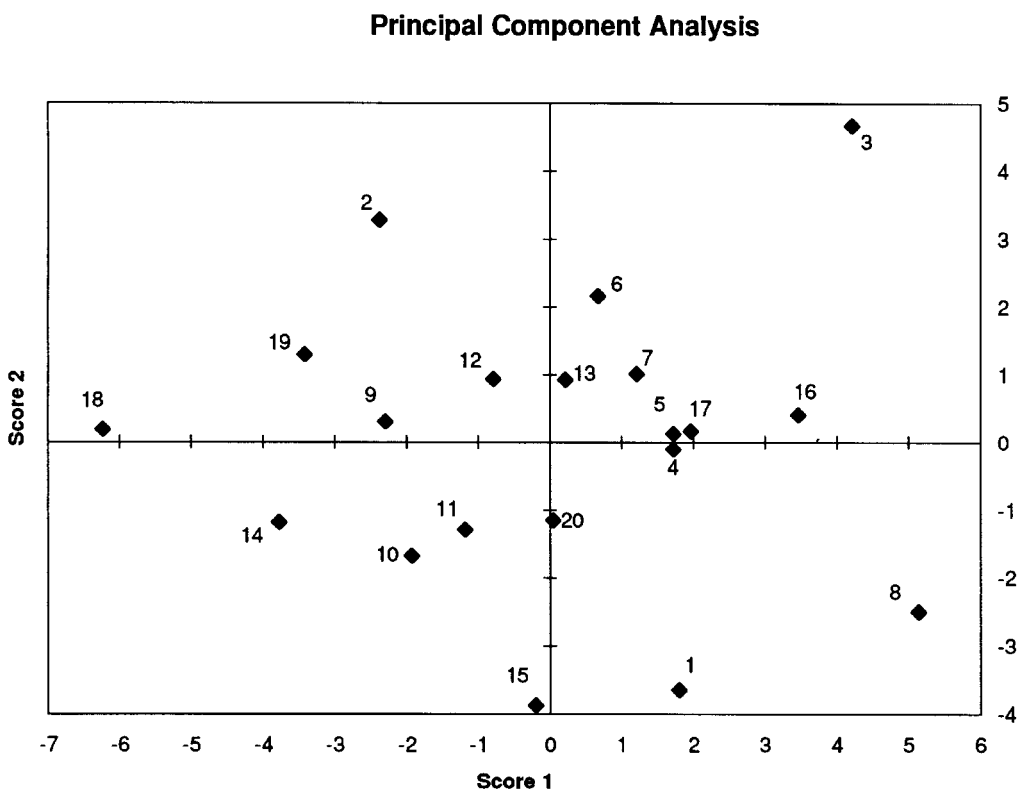

FIGURE 4. The first two scores from the principal component analysis of the MolSurf descriptors for the 20 coded amino acids.

Because the calculated MolSurf descriptors seem to be relatively insensitive to low energy conformational behavior, we used the MolSurf descriptors of the lowest found energy conformation as variables in the QSAR analyses of two dipeptide data sets.

The PCA on the MolSurf descriptor matrix resulted in four components that explained 66.3% of the variance in the matrix (29.8, 16.8, 10.0, and 9.7% for the individual components). Plots of the loadings and scores for the first two components are depicted in Figures 3 and 4, respectively. The first component (loading) is mostly influenced by global molecular parameters of the amino acid, while the second component is dominated by variables related to the nitrogen atom of the amino group as well as by side-chain related descriptors.

ACE INHIBITORS

A set of 58 dipeptides previously investigated by both Hellberg et al. using z scales[20] as well as by Collantes and Dunn using ECI/ISA parameters[25] was studied using the MolSurf descriptors. The same training set of nine peptides (VW, RW, YA, AA, FR, VG, GI, GR, DG) as selected by Hellberg et al. was used.[20] The PLS analysis resulted in two significant PLS components with $R^2 = 0.98$, $Q^2 = 0.77$, and root mean square error (RMSE) = 0.177. The model predicted the remaining 49 dipeptides (test set) with an RMSE of 0.554. A simple variable selection was performed using a LOO approach where each variable was left out from the model and its importance on predictivity, as judged by a LOO cross-validation procedure,[19] of the training set was assessed. If the predictivity of the model increased then the variable in question was permanently removed from the model, and otherwise the variable was permanently kept in the model. This resulted in a PLS model with three significant components and $R^2 = 0.99$, $Q^2 = 0.98$, and RMSE = 0.057 with 44 variables retained. The model predicted the remaining dipeptides with an RMSE of 0.483. The MolSurf based models

TABLE IV.
15 Most Important MolSurf Descriptors of PLS QSAR Model for ACE Inhibition Dipeptide Data Set.

PLS Regr. Coeff.	Pos	Descriptor	Type	Squared Term
− 0.0533	1	Polarity	m	sq
− 0.0532	1	Polarity	m	
0.0514	2	LA	s	sq
0.0478	1	CT	αC	
0.0467	2	I_{min}	αC	
− 0.0466	2	V_{max}	αC	
− 0.0465	2	I_{min}	αC	sq
− 0.0461	1	CT	αC	sq
− 0.0461	2	LB	s	sq
0.0461	2	log *P*	s	sq
0.0457	2	LA	s	
− 0.0436	2	V_{max}	αC	sq
0.0436	2	Polarizability	s	sq
0.0433	2	I_{min}	N	
− 0.0430	2	I_{min}	N	sq

Pos, position (1 or 2) in the dipeptide; m, the entire molecule (amino acid); s, amino acid side chain; αC, α-carbon; N, nitrogen atom of the amino group attached to the α-carbon atom.

TABLE V.
15 Most Important MolSurf Descriptors of PLS QSAR Model for Bitter Tasting Dipeptide Data Set.

PLS Regr. Coeff.	Pos	Descriptor	Type[b]	Squared Term
0.0768	2	log *P*	s	sq
0.0604	2	log *P*	s	
− 0.0538	2	I_{min}	αC	sq
0.0530	2	I_{min}	αC	
0.0516	2	I_{max}	N	
− 0.0513	2	I_{max}	N	sq
0.0482	2	Polarizability	s	sq
0.0480	1	log *P*	s	sq
0.0479	1	log *P*	s	sq
0.0442	2	Polarizability	s	
− 0.0440	2	Polarity	s	
− 0.0438	2	Polarity	s	sq
0.0437	2	Polarizability	m	
0.0436	2	log *P*	m	sq
− 0.0431	2	Polarity	m	

Pos, position (1 or 2) in the dipeptide; for types see Table IV for an explanation.

showed somewhat better predictivity compared with the corresponding models based on the *z* scales that gave RMSE values of 0.715 and 0.804 without and with variable selection, respectively, for the dipeptide test set.

Analysis of the PLS regression coefficients for the model based on all 58 ACE inhibitors (three significant PLS components with $R^2 = 0.87$, $Q^2 = 0.78$, and RMSE = 0.356) shows that the variables related to the second position have a larger impact compared to the first position. Out of the 15 most important variables, as judged by the scaled PLS regression coefficients (Table IV), 11 stems from position 2. From these (and other variables) one may conclude that the factors beneficial for high activity in position 2 are a polar amino acid residue with a polarizable and lipophilic side chain as well as the ability of the latter to act as a Lewis acid.

BITTER TASTING DIPEPTIDES

A PCA was performed on the bitter tasting dipeptide data set and five principal components were extracted that explained 66.6% of the variance (22.4, 18.9, 11.0, 7.6, and 6.7, respectively). A training set of 16 dipeptides was then selected based on the PCA scores weighted by their explained variances using a distance based design method.[26] This method uses a fast switching algorithm to maximize the minimum distance between two objects (in this case two dipeptides). The PLS analysis of the training set resulted in three significant PLS components with $R^2 = 0.98$, $Q^2 = 0.67$, and RMSE = 0.125. Using variable selection (see ACE Dipeptides for a description of the method), a PLS model with three significant PLS components and $R^2 = 0.97$, $Q^2 = 0.84$, and RMSE = 0.138 was derived. The predictivity of the test set for the two PLS models, as judged by the RMSE value, was 0.351 and 0.304, respectively. In this case, the corresponding models based on *z* scales (two significant PLS components) performed marginally better than the MolSurf models with RMSE values of 0.333, and 0.298, respectively, for the 32 dipeptides comprising the test set. An analysis of the PLS regression coefficients (see Table V for the 15 most significant coefficients) for the model based on all 48 dipeptides (three significant PLS components with $R^2 = 0.89$, $Q^2 = 0.79$, and RMSE = 0.209) shows that the second position influences the biological activity to a somewhat greater degree. Factors such as a polarizable as well as lipophilic and nonpolar side chain in both positions are favorable for high biological activity. This is manifested by WW, LW, FY, FW, and IW, which are the most active dipeptides.

RELATION BETWEEN AMINO ACID *z* SCALES AND Molsurf PARAMETERS

Because the PLS QSAR models yield statistics that are similar regardless of deployed amino acid residue descriptors for both peptide data sets investigated in this work, the obvious question with respect to the possible correlation between the two descriptor sets arises. The PLS analyses of each of the three z scales as dependent variable and the MolSurf descriptors resulted in two significant PLS components in all three cases. The ordinary and cross-validated squared correlation coefficients (R^2/cv $-$ R^2) for z scales 1−3 were 0.782/0.590, 0.853/0.697, and 0.723/0.228, respectively. The analyses reveal that the two first z scales are rather well described by the MolSurf parameters according to cross-validation while the third z scale is less well modeled by the MolSurf descriptors. Thus, the same kind of information is found in the first two, most important, z scales and the MolSurf parameters to a large degree while new information, not present in the MolSurf descriptors, is found in the third, less important, z scale. Therefore, a rather high degree of correlation with respect to information regarding amino acid characterization is found between the z scales of Hellberg et al.[20] and the MolSurf descriptors derived and presented in this work. This, in turn, explains the rather similar statistical outcome of the PLS analyses of the two peptide data sets using the z scales and MolSurf descriptors, respectively.

Conclusions

This work shows that MolSurf parameters, which are relatively insensitive to conformational changes and easy to compute, can be used to derive useful descriptors for amino acids. These descriptors can subsequently be used in deriving QSARs for peptides. The MolSurf descriptors are not necessarily superior to other parameters but result in statistically significant QSAR models. A notable correlation exists between the z scales of Hellberg et al.[20] and the MolSurf descriptors. The MolSurf descriptors are easy to interpret and are thus helpful when designing new peptides with improved properties.

References

1. C. Hansch and A. J. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979.
2. C. G. Swain and E. C. Lupton, *J. Am. Chem. Soc.*, **90**, 4328 (1968).
3. *MolSurf Version 1.1*, Qemist AB, Karlskoga, Sweden, e-mail: par.sjoberg@mbox309.swipnet.se.
4. Spartan Version 4.1, Wavefunction, Inc., Irvine, CA.
5. Macromodel Version 5.5, Dept. Chem., Columbia Univ., New York.
6. T. Brinck, J. S. Murray, and P. Politzer, *J. Org. Chem.*, **58**, 7070 (1993).
7. J. S. Murray, T. Brinck, and P. Politzer, *J. Phys. Chem.*, **97**, 13807 (1993).
8. M. Haeberlein and T. Brinck, *J. Chem. Soc. Perkin Trans.* **2**, 289 (1997).
9. T. Brinck, J. S. Murray, P. Politzer, and R. E. Carter, *J. Org. Chem.*, **56**, 2934 (1993).
10. H. Hagelin, J. S. Murray, T. Brinck, M. Berthelot, and P. Politzer, *Can. J. Chem.*, **73**, 483 (1995).
11. T. Brinck, J. S. Murray, and P. Politzer, *Mol. Phys.*, **76**, 609 (1992).
12. J. S. Murray, T. Brinck, P. Lane, K. Paulsen, and P. Politzer, *J. Mol. Struct.* (*Theochem.*), **307**, 55 (1994).
13. P. Sjöberg, In *Computer-Assisted Lead Finding and Optimization: Current Tools for Medicinal Chemistry*, H. van der Waterbeemd, B. Testa, and G. Folkers, Eds. VHCA, Basel, Switzerland, 1997, p. 83.
14. J. E. Jackson, *A Users Guide to Principal Components*, Wiley, New York, 1991.
15. I. T. Jolliffe, *Principal Component Analysis*, Springer−Verlag, New York, 1986.
16. E. R. Malinowski, *Factor Analysis in Chemistry*, 2nd ed., Wiley, New York, 1991.
17. S. Wold, E. Johansson, and M. Cocchi, In *3D QSAR in Drug Design*, H. Kubinyi, Ed., ESCOM, Leiden, The Netherlands, 1993, p. 523.
18. S. Hellberg, L. Eriksson, J. Jonsson, F. Lindgren, M Sjöström, B. Skagerberg, S. Wold, and P. Andrews, *Int. J. Pept. Protein Res.*, **37**, 414 (1991).
19. S. Wold, *Technometrics*, **20**, 379 (1979).
20. S. Hellberg, M. Sjöström, B. Skagerberg, and S. Wold, *J. Med. Chem.*, **30**, 1126 (1987).
21. U. Norinder, unpublished results.
22. H. Kubinyi, *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM, Leiden, The Netherlands, 1993.
23. K. H. Kim and Y. C. Martin, *J. Med. Chem.*, **34**, 2056 (1991).
24. K. H. Kim, *Quantum Struct.-Act. Rel.*, **12**, 232 (1993).
25. E. R. Collantes and W. J. Dunn III, *J. Med. Chem.*, **37**, 2705 (1995).
26. E. Marengo and R. Todeschini, *Chem. Intel. Lab. Syst.*, **16**, 37 (1992).